

HOWTO 20 (Workbench 1): Tekst analyseren en Grammaticale Info raadplegen (POS-tagging)

Deze handleiding legt uit hoe je een tekst in de **Workbench** plaatst, deze laat analyseren door de computer (Part-of-Speech tagging = toekennen van een woordsoort aan elk woord) en hoe je de gedetailleerde grammaticale informatie per woord raadpleegt.

1. Wat doet de Analyse-functie?

De analyse-functie van de Lexogoth Workbench:

- **Identificeert woordsoorten:** Herkent of een woord een zelfstandig naamwoord, werkwoord, adjectief, etc. is.
- **Bepaalt de lemma's:** Koppelt vervoegde of verbogen woorden aan hun grondvorm (bijv. *mangeons* → *manger*).
- **Biedt contextuele hulp:** Geeft directe toegang tot grammaticale voetnoten en uitgebreide vervoegingen.

De plus, l'Organisation mondiale de la santé (OMS) a annoncé jeudi que l'Argentine envoyait vers cinq pays quelque 2500 kits pour dépister l'hantavirus, objet d'une alerte sanitaire internationale depuis la découverte d'un foyer sur un navire de croisière. "Nous avons [...] organisé l'envoi de 2500 kits de diagnostic depuis l'Argentine vers les laboratoires de cinq pays", a annoncé le chef de l'OMS, Tedros Adhanom Ghebreyesus, aux journalistes à Genève.

L'OMS a également "informé les 12 pays dont des ressortissants ont débarqué à Sainte-Hélène. Ces 12 pays sont le Canada, le Danemark, l'Allemagne, les Pays-Bas, la Nouvelle-Zélande, Saint-Kitts-et-Nevis, Singapour, la Suède, la Suisse, la Turquie, le Royaume-Uni et les États-Unis", a déclaré devant la presse à Genève Tedros Adhanom Ghebreyesus, directeur général de l'OMS.

The screenshot shows the Lexogoth Workbench interface. At the top, there is a blue header with the text 'GEGEVENS'. Below this, there is a search bar with the text 'Aangeklikt w' and a search button labeled 'OPZOEKEN LEXOGOTH-CORPUS'. The main content area shows the word 'de plus' being analyzed. Below the word, there is a table with the following columns: 'Lemma' (de plus), 'Woordsoort' (MWE_PART), and 'MWE_PART'. Below the table, there is a section titled 'UITDRUKKING (MWE)' with the text 'Vaste vorm de plus'. Below this, there is a section titled 'COMBINATIES (Q-tags)' with the text 'Basisvorm de plus'. At the bottom, there are two tabs: 'Annotaties' and 'Grammatica'.

2. Stap 1: Tekst invoeren

Voordat je kunt analyseren, moet er tekst in de Workbench staan.

1. Kopieer een tekst uit een extern document (Word, PDF, website).
2. Klik in het grote witte tekstveld in de Workbench waar de placeholder "Voer hier de Franse tekst in..." staat.
3. Plak de tekst (**Ctrl + V**) of gebruik het contextmenu met de muis. Merk op dat de tekst automatisch gefilterd wordt van foto's en andere niet-talige informatie die typisch is voor html-tekst.

3. Stap 2: De Analyse uitvoeren

Zodra de tekst in het venster staat, moet de "motor" van de Workbench de tekst verwerken.

1. Klik in de bovenste menubalk op de knop **Analyseer** (naast het groene herlaad-icoon).
2. De tekst wordt nu "getagd": de computer loopt elk woord na en kent er grammaticale labels aan toe.
3. **Resultaat**: woorden worden interactieve elementen waar je op kunt klikken. De informatie kan je onderaan in de tab "Grammatica" uitlezen. Om voetnoten toe te voegen of te lezen kies je voor de tab *Annotaties*.

4. Stap 3: Woorden onderzoeken

Na de analyse kun je dieper in de tekst duiken:

1. **Woord aanklikken**: Klik op een willekeurig woord in je geanalyseerde tekst.
2. **Grammaticavenster**: Onderaan de interface verschijnt nu in het tabblad **Grammatica** de specifieke info:
 - **Lemma**: De basisvorm van het woord.
 - **POS-tag**: De grammaticale categorie (bijv. *V* voor werkwoord, *NC* voor substantief).
 - **Kenmerken**: Info over getal, geslacht of tijdstip.
3. **Voetnoten**: Als er bij een woord een specifiek grammaticaal aandachtspunt hoort, verschijnt dit als een verklarende voetnoot in dit onderste venster.
4. Voor sommige woorden verschijnt er ook **MWE** (uitdrukking van meerdere woorden) of **Q-tag** (nuttige woordcombinaties, vaste uitdrukkingen,...)

5. Stap 4: Geavanceerd zoeken via het Grammaticavenster

Wil je weten of een specifiek woord of patroon vaker voorkomt in je database?

1. Selecteer een woord in de tekst.
2. In het **Grammaticavenster** (onderaan) zie je een icoontje van een **vergrootglas**.
3. Klik op dit vergrootglas: de Workbench voert direct een zoekopdracht uit in de Lexogoth-database naar alle zinnen die dit specifieke woord (of deze grammaticale vorm) bevatten.

6. Tips voor een scherpe analyse

- **Sequentie taggen:** Je kunt ook een specifiek tekstfragment selecteren en alleen die sequentie opnieuw laten taggen als je wijzigingen hebt aangebracht.
- **Kleurgebruik:** Let op de kleurveranderingen van de knop *Analyseer*; indien deze op oranje staat, betekent dit dat de tekst sinds de laatste analyse veranderde en de tags en indexen mogelijk niet meer kloppen. Als dit het geval is, druk dan terug op *Analyseer > Tekst volledig taggen en indexeren* om de analyse up-to-date te houden.
- **Correctie:** Klopt een tag niet? In de 'Bewerken'-modus kun je vaak handmatig de info in het grammaticavenster bijsturen. Let op deze informatie gaat wel verloren bij het hertaggen van de tekst!